Blind deconvolution via independent component analysis for thin-pavement thickness estimation using GPR

Khaled CHAHINE¹, Vincent BALTAZART¹, Yide WANG², and Xavier DÉROBERT¹

¹Laboratoire Central des Ponts et Chaussées, Nantes, France, khaled.chahine@lcpc.fr
²Ecole polytechnique de l’université de Nantes, Nantes, France

Abstract

Within the scope of GPR application to pavement survey, the sparse reflectivity series representing the layered structure of the pavement is convolved with the radar wavelet. A successful blind deconvolution would then retrieve the latter reflectivity series and improves the time resolution without any a priori information. In this paper, we cast the convolutional model as a multidimensional data model which renders blind deconvolution via independent component analysis (ICA) possible. We use a nonlinear contrast function which is matched to the sparse nature of the reflectivity series. The method is tested on synthetic and real GPR data from a thin PVC slab. The results attest to the accuracy of the time delay estimates and verify the high resolution of the proposed approach.

Keywords

Nondestructive testing, reflectivity series, sparsity, time delay estimation, time resolution

1 Introduction

In the field of civil engineering, ground penetrating radar is used as a nondestructive testing technique. Being characterized by wavelengths of the order of tens of centimeters, it is more dedicated to pavement survey applications. A short-duration electromagnetic wave, the radar wavelet, is transmitted into the ground and the reflected wave, the radar trace, resulting from dielectric contrasts is measured. Under simplifying assumptions, such trace can be modeled as the convolution between the reflectivity series of the medium and an unknown wavelet [1].

In comparison with conventional deconvolution (e.g. see [4,5]), we cast the convolutional model as a multidimensional data model which renders blind deconvolution via independent component analysis (ICA) possible. In comparison with [6], another contrast function is introduced that encourages the sparsity of the reflectivity series corresponding to a stratified-earth model. The estimated reflectivity series indicates the position in time of layers, and thus provides the time delay estimate associated with each interface. ICA and other blind source
separation techniques have already been applied to GPR and seismic data to reduce clutter and suppress multiple reflections.

2 Data model

The convolutional model, widely used in the seismic domain, is an accepted model for the description of a radar trace. In this model, the trace can be expressed as the convolution of two signals; a radar wavelet and the reflectivity series:

\[ x(t) = a(t) \ast s(t) \]  

(1)

where \( a(t) \) is the radar wavelet, \( s(t) \) is the reflectivity series, and \( \ast \) is the convolution operator.

The reflectivity is a sequence of spikes that indicates the position in time of layers in the subsurface. It permits, if successfully retrieved, to estimate the corresponding thicknesses. It is clear from the above equation that we are dealing with one equation and two unknowns. Generally, such a problem is tackled using blind deconvolution. In fact, a close affinity can be found between blind deconvolution and ICA (more precisely, the special case of blind deconvolution where the considered signal samples are independent and identically distributed over time) [7].

The discrete convolutional model is given by:

\[ x(t_j) = \sum_j a(t_{j-j+1}) s(t_j) \]  

(2)

which can be also expressed in matrix notation as:

\[ x = As \]  

(3)

where

\[ x = [x(t_1), x(t_2), \ldots, x(t_n)]^T \]

\[ s = [s(t_1), s(t_2), \ldots, s(t_n)]^T \]  

(4)

To express the convolution operation under matrix notation, the mixing matrix is constructed from delayed versions of the radar wavelet as follows:

\[ A = [N_a, N_2a, \ldots, N_na] \]  

(5)

where

\[ a = [a(t_1), a(t_2), \ldots, a(t_m)]^T \]  

and \( N_i = \{0_{i-1}, 1_{m-n}, 0_{(n-m-i+1)}\} \) is a zero padding matrix, \( 0 \) is the zero matrix, \( I \) is the identity matrix, and \( i = 1, 2, \ldots, n \).

Although this model can be recognized as an ICA model, it provides only one realization which is inadequate for ICA [8]. To overcome this difficulty, we consider delayed versions of the radar trace and the reflectivity series:

\[ S = [z^{n-1}s, z^{n-2}s, \ldots, zs, s]^T, \]

\[ X = [z^{n-1}x, z^{n-2}x, \ldots, zx, x]^T \]  

(6)

where \( z \) is the unit time delay. After this reorganization, the \( j^{th} \) realization (row or column) of \( X \) becomes the convolution of the \( j^{th} \) realization of \( S \) with the radar wavelet. Now that a multidimensional data model is available, the blind deconvolution problem is posed in a way that can be solved by ICA.
3 Independent component analysis

ICA aims at recovering latent independent sources from their observable linear mixtures [9]. Denote by \( \mathbf{x} = (x_1, \ldots, x_n)^T \) the vector of observable signals, where \( T \) is the transpose operator. \( \mathbf{x} \) is assumed to be generated by \( \mathbf{x} = \mathbf{As} \) where \( \mathbf{s} = (s_1, \ldots, s_n)^T \) is of mutually independent components and \( \mathbf{A} \) is the mixing matrix. Without any a priori information about the mixing process or the sources, the objective of ICA is to find a linear transformation \( \mathbf{W} \) such that the retrieved components \( \mathbf{y} = \mathbf{Wy} \) are as independent as possible in the sense of optimizing some function \( I(y_1, \ldots, y_n) \) that measures independence. From [10], \( I \) is defined as the mutual information among the \( n \) scalar random variables \( y_i, i = 1, \ldots, n \). Expressing mutual information in terms of negentropy gives:

\[
I(y_1, \ldots, y_n) = J(y) - \sum_i J(y_i)
\]

(7)

Mutual information is a natural measure of the independence of random variables and it is reasonable to use it as the criterion for finding the ICA transform \( \mathbf{W} \). Thus we define in this paper, following [10], the ICA of a random vector \( \mathbf{x} \) as an invertible transformation \( \mathbf{y} = \mathbf{Wy} \) where the matrix \( \mathbf{W} \) is determined so that the mutual information of the transformed components \( y_i \) is minimized which is equivalent to finding directions in which negentropy is maximized. The mutual information is not affected by the multiplication of the components by scalar constants. Therefore, this definition only retrieves the independent components up to some multiplicative constants.

The identifiability of the ICA model can be ensured by imposing the following restrictions:

- All the latent sources, with the possible exception of one, must be non-Gaussian.
- The number of observed linear mixtures must be at least equal to the number of latent sources.
- The mixing matrix must be of full column rank.

4 Contrast function and sparsity

To use the aforementioned definition of ICA, a simple approximation of the negentropy is needed. Employing the approximation in [14] the expression of negentropy becomes:

\[
J(y_i) \approx \left[ E \{G(y_i)\} - E \{G(\nu)\} \right]^2
\]

(8)

where \( G \), the contrast function, is practically any non-quadratic function and \( \nu \) is a Gaussian variable of zero mean and unit variance. The question becomes: how to choose the contrast function so that sparsity of the components is encouraged?

In [11], the equivalence between a sparse signal and a super-Gaussian signal is suggested. Based on the a priori information provided by the nature of the reflectivity series, we consider the unknowns as realizations of the following super-Gaussian density called the double exponential density:

\[
f(s) = k_1 \exp(-k_2|s|)
\]

(9)

where \( k_1 \) and \( k_2 \) are normalization constants that ensure that \( f(s) \) is a density function. The heavy-tailed nature of this density function yields further insight into the sparse nature of the solutions. The corresponding optimal contrast function for estimating an independent component whose density function is \( f(s) \) takes the form:
where the constants have been dropped for simplicity. However, the problem with such a contrast function is that it is not differentiable at 0. An approximating differentiable function having the same qualitative behavior is given by \( G(s) = \log(\cosh(\alpha s)) \), \( \alpha \geq 1 \) is a constant.

The nonlinearity which is the derivative of the approximating function is then the familiar \text{tanh} function. This nonlinearity is compatible with the sparse nature of the reflectivity series.

### 5 Computer simulations

A simplified two-layered pavement is considered in this section to make the problem more tractable. Besides, we restrict ourselves to the air-coupled Radar scenario. For this noiseless simulation, the radar wavelet is a 900 MHz Ricker wavelet of 128 samples. This type of wavelets is widely accepted in the literature to approximate the source waveform of some commercially available pulsed GPR systems. A reflectivity series consisting of three spikes representing the surface echo, the base-layer echo, and the echo from an assumed metallic plate is considered.

These spikes indicate arrival times of 0.9337 ns, 1.228 ns and 2.848 ns, respectively. So, the differential time delay of the first couple of echoes is 0.2945 ns and that of the second is 1.62 ns. At a 900 MHz frequency, the Ricker wavelet has a duration of approximately 1.1 ns (GPR systems are designed to obey the following relation between central frequency and bandwidth \( B \approx f_c \text{ where } B = 1/T \text{ and } T \text{ is the wavelet duration} \)). Therefore, by convolving this wavelet with the generated reflectivity series, we simulate a scenario of overlapping echoes. From Fig.1 we can distinguish two distinct wavelets while in fact there are three, the first two being confounded. This is due to the fact that for a Ricker wavelet whose central frequency is almost equal to the bandwidth, the smallest detectable thickness is:

\[
e_{\text{min}} \geq \frac{\lambda_{\text{medium}}}{2} = \frac{\lambda_{\text{air}}}{2\sqrt{\varepsilon'}}
\]

where \( \lambda_{\text{medium}} \) and \( \lambda_{\text{air}} \) are the wavelengths in the medium and air, respectively. In our case, this thickness is 9.6 cm for \( \varepsilon' = 3 \) while we are trying to detect a thickness of almost 2.6 cm corresponding to the 0.2945 ns time delay. In such a case, the needed time resolution is \( B \times 0.2945 \approx 0.27 < 1 \).

\[\text{Figure 1. The received trace with only two discernible reflections.}\]

\[\text{Figure 2. The retrieved reflectivity series by ICA. Asterisks designate the true locations of spikes.}\]
series. To choose the best candidate, we subtract the convolution of each couple (a couple is an independent component and its corresponding row in the estimated mixing matrix) from the radar trace. The independent component of the couple yielding the smallest value is retained as the best estimate of the reflectivity series. The one corresponding to our simulation is shown in Fig.2. From Fig.2, it is obvious that the algorithm is capable to distinguish between overlapping echoes and thus enhances the time resolution of GPR. The linear phase shift of the retrieved series is due to how the mixtures in $X$ are organized, and so the true reflectivity had to be shifted in order to carry out the comparison.

6 Experimental test

To test the performance of the proposed approach on real data, we used far-field measurements on a 2-cm-thick polyvinyl chloride (PVC) slab set on a metallic plate. The dielectric constant of the slab at 2 GHz was found to be $2.9 + j0.015$ [3]. These measurements were conducted by [3] using a GPR (GSSI SIR-3000) combined with a bistatic device. The central frequency is 1.7 GHz in air, and the bandwidth is equal to about 2 GHz at -6 dB. Consequently, the proposed approach is faced with a setup of overlapping echoes of time resolution $B\Delta \tau = 2 \times 0.22706 \approx 0.45 < 1$, where $B$ is the bandwidth and $\Delta \tau$ is the time delay of backscattered echoes. The acquired radar trace consists of 1024 samples and has a 10 ns time range. Fig.3 shows the raw experimental data where the first echo is the airwave between the transmitter and receiver while the overlapping second and third echoes correspond to the backscattered echoes from the two interfaces.

![Figure 3. The received radar trace. The rectangle indicates the zone of overlapping echoes.](image)

![Figure 4. The time-filtered radar trace](image)

![Figure 5. The deconvolved reflectivity series.](image)

The trace in Fig.3 is time filtered as shown in Fig.4 so as to keep the echoes from these interfaces and eliminate the airwave and any multiples. SNR is about 35 dB on the latter trace. Fig.5 demonstrates the good performance of the method and that it satisfies a condition of sparsity [2]. The time delay between the retrieved spikes is 0.225 ns which corresponds to a thickness of 1.9819 cm with a relative error of 0.9 %. Note that the absolute locations of the
spikes do not necessarily correspond to the true ones; however, the time delay between them is accurate.

7 Conclusions

We investigated the feasibility of blind deconvolution via ICA for the problem of thin-pavement thickness estimation. We used a nonlinearity which is compatible with the sparse nature of the reflectivity series with the aim of enhancing the time resolution of GPR. The proposed approach was tested on synthetic and real data. In both cases, the results were satisfactory.

Acknowledgements

The authors would like to thank the region “Pays de la Loire” which partially funds this research work.

References